Project Title: "Complete genome sequencing of *Shewanella putrefaciens (oneidensis)* MR-1".
DOE Award Number: DOE DE-FG02-97ER62444
Principle Investigator: John F. Heidelberg
Recipient: The Institute for Genomic Research
Address: 9712 Medical Center Drive, Rockville, MD 20850
FINAL REPORT

Because of the genome size of *Shewanella putrefaciens (oneidensis)* (predicted as 4.2 Mbp) 70% of the cost for genome sequencing were requested from grant DOE DE-FG02-97ER62444. These funds were expected to allow completion of the low-pass (5-fold) random sequencing and complete closure and annotation of the 200 kbp plasmid. Because of cost reduction that occurred during the period of this grant we have been able to far exceed these goals. Currently, the *S. putrefaciens (oneidensis)* MR-1 is very nearly completely closed, even though the genome was significantly larger than expected (5.1 Mbp or 18% larger than expected) and extremely repetitive. The entire genome sequence has been made BLAST searchable on the TIGR web page and an extensive effort has been made to make data and analyses available to all researchers working on *S. putrefaciens (oneidensis)* MR-1. Because additional funds were made available, the closure-sequencing phase on this bacterium is continuing, and is expected to be closed by the end of this year. Below is the scope of work preformed on Award DOE DE-FG02-97ER62444:

Library construction and random sequencing phases of this project have been competed. Genome closure phase is 99% complete. Preliminary genome assembly analysis to monitor the progress of the genome closure phase, and to identify sequencing areas in need of special attention have been performed.

Random Sequencing Phase.

Plasmid sequencing. Shewanella *putrefaciens (oneidensis)* MR-1genomic DNA was shotgun cloned into pUC18 to construct a random genomic library. At the end on the random sequencing phase of this project 67,815 plasmid sequences had been obtained. Average sequence length of plasmids done in the random phase was 527 nucleotides.

Lambda sequencing. To aid in the genome closure phase of S. *putrefaciens (oneidensis)* MR-1, a total 635

lambda sequences were obtained from a lambda library containing 18 - 20 kb inserts.

Genome Closure Phase.

DISCLAIMER

At the completion of the random sequencing phase of the S. *putrefaciens (oneidensis)* MR-1project all sequences were assembled into contigs with the TIGR_assembler program, and research associates edited the contigs. After the sequence editing, TIGR_assembler was rerun and the contigs were linked to one another by shared common clones. These linked sets of contigs are referred to as groups.

After the most resent assembly and grouping, (completed August 2000) there are 11 over 1,500 bp (193 physical gaps were closed in the last year), and two of the remaining assemblies are closed plasmids. The largest single assembly is 2,052,242. The entire genome is currently linked by sequence information.

Physical gaps. We closed the physical gap several strategies were employed including, direct sequencing genomic DNA, multiplex PCR, and Genome Walker (CloneTech).

Repeats. Repeated sequences were identified with the repeat_finder.spl program. The repeats were categorized in three types, 1) repeated sequence was smaller then the read length (< 500 bp), 2) repeat is larger than a sequence read, but smaller than the clone insert size (500-3,000 bp) and 3) repeat is larger then the clone insert. Nothing is done with the first sized repeat. The size 2 repeats have a clone sequenced from unique flank to unique flank sequence. The size 3 repeat has either a lambda clone walked if one is available or a PCR product produced spans the repeat which is subsequently walked. The S. *putrefaciens (oneidensis)* MR-1genome has more repeated sequences than any other competed genome, with over 170 transposons and 9 rRNA operons. All the transposons have been confirmed and 7 of the 9 rRNA operons are competed.

Ribosomal RNA operons are a special repeat that have presented a complication in genome closure for some previous genomes, including *Haemophilus influenzae* and *Vibrio cholerae*. There are two reasons for this difficulty: first, in organisms that have multiple copies of the rRNA operon, this operon represents a large repetitive region which can be difficult to PCR; second, in organisms that have a high sequence similarity to *E. coli* the rRNA promoters are unclonable. We estimate that S. *putrefaciens (oneidensis)* MR-1has 9 separate rRNA operons and the promoter region was unclonable. We have identified the 16S rRNA flank sequence by genome walking, and have paired ends by combinatorial PCR from unique 16S rRNA flanking sequence to 5S rRNA flanking sequence.

Preliminary annotation.

Periodically, we have searched the S. *putrefaciens (oneidensis)* MR-1sequences against our in-house non-redundant database. These preliminary observations have show similarities between S. *putrefaciens (oneidensis)* MR-1and *V. cholerae* in some role categories (small molecule biosynthesis, central intermediary metabolism) but differences in others (sugar metabolism). It will be interesting to examine these similarities and differences in light of the different ecological niches occupied by these two organisms. We has also searched for genes of special interest to the Nealson laboratory (cytochromes, ferridoxins, and sulfate reductases) and made a complete list of the preliminary open reading frames available to the Zhou laboratory.